Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Harris Hawks optimization-based cellular automata model for urban growth simulation

Yuan Ding

dingyuanhhu@hhu.educn

Hohai University

Hengyi Zheng

Hohai University

Fuming Jin

Hohai University

Dongming Chen

Xinyu Huang

Research Article

Keywords: urban growth, Harris Hawks optimization, cellular automata, transition rules mining, spatial dynamic simulation

Posted Date: June 26th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4636601/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Harris Hawks optimization-based cellular automata model for urban growth simulation

Yuan Ding(corresponding author)[1], Hengyi Zheng[2], Fuming Jin[2], Dongming Chen[2], Xinyu Huang[2]

[1] College of Geography and Remote Sensing，Hohai University，Nanjing, China

[2] School of Earth Science and Engineering，Hohai University，Nanjing, China

E-mail Address of the Corresponding Author: dingyuanhhu@hhu.edu.cn

ABSTRACT:

This paper proposes an innovative cellular automata model based on the Harris Hawk Optimization (HHO) algorithm. HHO is an intelligent optimization algorithm inspired by the cooperative hunting behavior of Harris's hawks, demonstrating excellent optimization efficiency in spatial searches. Combining the HHO algorithm with the CA model, we establish the HHO-CA model for simulating urban growth in Guangzhou, China. The simulation achieves a total accuracy of 91.95%, an accuracy of urban cells of 82.43%, and a Kappa coefficient of 0.7441, all superior to the Null model. Furthermore, comparing the HHO-CA model with other representative CA models, the HHO-CA model outperforms in total accuracy, accuracy of urban cells, and Kappa coefficient, showcasing significant advantages in using the HHO algorithm to mine transition rules during the simulation of urban growth processes.

Keywords: urban growth; Harris Hawks optimization; cellular automata; transition rules mining; spatial dynamic simulation

1. Introduction

In recent decades, with the accelerated pace of urbanization, urban growth and sprawl have been identified as the most crucial aspects of urbanization and urban development(Zhong et al., 2023; Wei and Ewing, 2018). The disorderly expansion of cities is expected to exert greater pressure on issues such as urban land resource management and spatial allocation(Wei and Ewing, 2018; Huang et al., 2020). Urban

growth and the various problems it triggers have long been a focal point for many scholars(Song et al., 2020; Xie et al., 2020). The close correlation between urban growth and multiple driving factors has been a research hotspot(Salem et al., 2021; Rubiera-Morollón and Garrido-Yserte, 2020; Msofe et al., 2019). In the simulation studies of urban growth, common models include Support Vector Machine (SVM)(Karimi et al., 2019), Markov Chain(Hamdy et al., 2016), and Logistic Regression(Tayyebi et al., 2010). Among them, cellular automata (CA) models have demonstrated remarkable capabilities in simulating urban growth(Chen et al., 2020).

The CA model is renowned for its powerful ability to simulate the spatiotemporal evolution of complex geographical phenomena(Cai and Wang, 2020). Faced with intricate global spatial patterns, the CA model can be composed of simple rules, making this bottom-up approach more suitable for explaining complex theories(Liu et al., 2008). CA models have been widely applied in studies related to land-use change simulation(White and Engelen, 1993; Clarke and Gaydos, 1998; Li and Yeh, 2002; Maria de Almeida et al., 2003; Wu, 2002), urban growth(Zhou et al., 2022; Batty et al., 1999; He et al., 2008; Li et al., 2014; Couclelis, 1997), among others. Existing research has indicated that CA models are highly suitable for simulating the processes of geographical evolution.

The definition of transition rules is the core of CA models(Guan et al., 2016). Currently, various optimization algorithms have been employed for mining CA models transition rules, such as particle swarm optimization (PSO)(Lu et al., 2022), whale optimization algorithm (WOA)(Ding et al., 2022), cuckoo search (CS)(Cao et al., 2015), and artificial fish swarm algorithm (AFSA)(Huang et al., 2021). The use of swarm intelligence optimization algorithms has proven feasible for mining CA transition rules, contributing significantly to a better understanding of model transition rules.

In this study, the Harris Hawks optimization (HHO) algorithm(Heidari et al., 2019) is utilized to mine CA transition rules. HHO is a novel metaheuristic intelligence algorithm proposed in recent years, simulating the cooperative hunting behavior of Harris's hawks. The research will construct the HHO-CA model for simulating and optimizing urban growth, applying it to the dynamic simulation of urban growth in

Guangzhou city. A quantitative comparison with the NULL model will be conducted to analyze the practicality and effectiveness of the HHO-CA model.

The remaining sections of the article are organized as follows. Section 2 explains the details of the proposed HHO-CA model. Section 3 introduces the study area and the experimental dataset. Section 4 discusses the experimental results. Section 5 provides a summary of the article and outlines prospects for future research.

2. Methods

2.1 Rule construction

The transition rule is a key element of study in CA models, describing the mapping relationship between the current state of a cell and its state at the next moment. In other words, the transition rule refers to a functional mapping, which is formally expressed as follows:

$$S_i^{t+1} = f(S_i^t, N_i^t) \quad (1)$$

Where S_i^{t+1} represents the state of cell i at time $t+1$, S_i^t represents the state of cell i at time t, N_i^t represents the neighborhood attributes of cell i at time, and f represents the transition rule.

In the context of geographic cellular automata (GCA), each cell status is composed of specific attribute conditions. Each feature attribute determining the cell status has upper and lower bounds, these values define the possible range of values for that attribute. Within each attribute interval, feature attribute nodes are selected and connected to form combinations. Finally, a cell status is determined, serving as the state of the cell in the next time step that satisfies all conditions in the combination. This completes the construction of the transition rules. Additionally, attribute values should be appropriately normalized to a consistent range before execution. In this research, all attribute values, whether continuous or discrete, are normalized to the [0, 1] range. By utilizing the HHO algorithm to find optimal thresholds for attributes, the best connections between attribute conditions and cell status can be identified. As illustrated

in the following **Fig 1**, each connection is related to the transition rules of the CA model. The process of mining transition rules can be seen as finding the optimal links between attribute conditions and cell status, specifically seeking the optimal upper and lower bounds for attribute conditions.

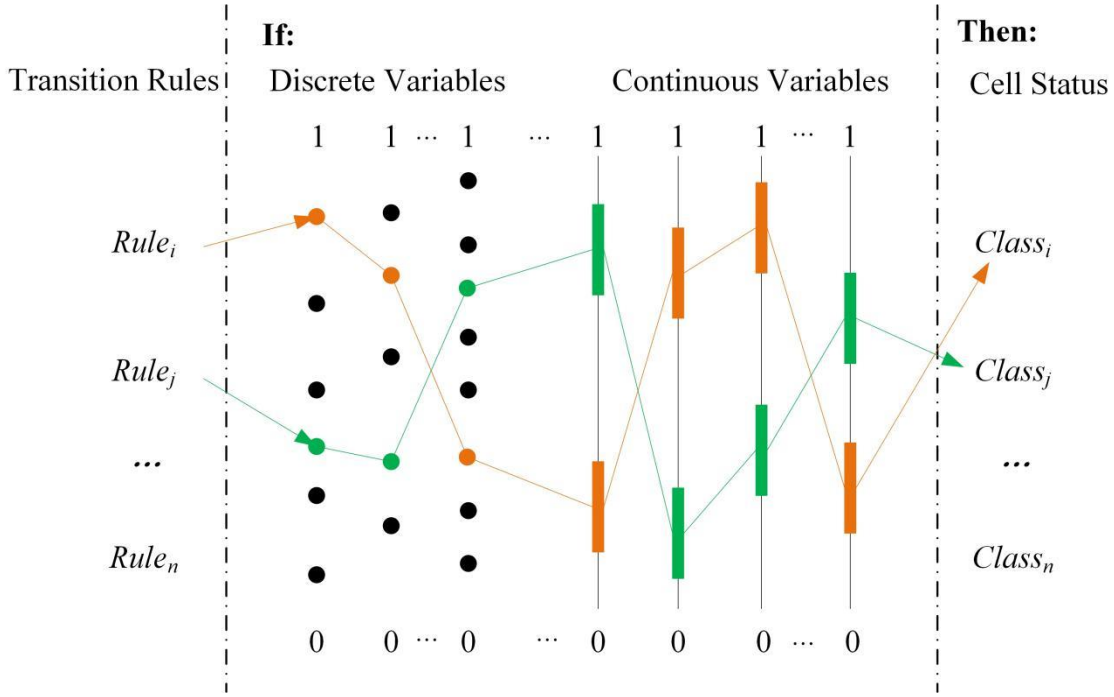

Fig 1. Schematic of transition rules based on conditional statements

The transition rules of the CA model can be expressed using the logical statement "**If...Then...**", which is straightforward for understanding. The specific expression is as follows:

If
attribute1 ≥ lower_1 and attribute1 ≤ upper_1
and
attribute2 ≥ lower_2 and attribute2 ≤ upper_2
and
... ...
and
attribute_n ≥ lower_n and attribute_n ≤ upper_n
Then
cell status = $class_i$

2.2 HHO for mining transition rules

2.2.1 The Harris Hawk optimization

The Harris Hawks optimization (HHO) algorithm is a metaheuristic intelligence algorithm proposed by AA Heidari, S. Mirjalili, H. Faris, and others in 2019. The unique group hunting behavior of Harris's hawks is well-suited to be simulated as a swarm intelligence optimization process. The HHO algorithm can continuously search for the optimal solution in the search space, exhibiting excellent optimization efficiency and convergence accuracy.

In the process of utilizing the HHO algorithm for mining CA transition rules, the initial vector elements' count is twice the number of feature attributes, due to the CA transition rules being composed of the minimum and maximum values of each attribute value. This is denoted as $X_i = (X_i^1, X_i^2, .., X_i^{2k-1}, X_i^{2k}, .., X_i^{2*dim-1}, X_i^{2*dim})$, where *dim* represents the number of cell feature attributes. The specific initialization method is expressed as Equation (2).

$$\begin{cases} X_i^{2K-1} = lb + rand(0,1)(ub - lb) \\ X_i^{2K} = X_i^{2K-1} + rand(0,1)(lb - X_I^{2K-1}) \end{cases} \quad (2)$$

Where X_i^{2k-1} indicates the lower bound of the value at the *K* attribute, X_i^{2k} represents the upper bound of the value at the *K* attribute; *i* denotes the individual; *ub* denotes the maximum value of the *K* attribute, *lb* denotes the minimum value of the *K* attribute, and *rand(0,1)* signifies that the range of the random variable is [0,1].

In the HHO algorithm, the behavior of Harris's hawks involves three steps: detecting prey, evaluating measures, and launching attacks.

(1) Detecting Prey. During this phase, Harris's hawks randomly search for prey (the optimal solution in the solution space), simulating the behavior of a hawk flock observing prey from different positions. They detect prey based on two strategies:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2X(t)|, q \geq 0.5 \\ [X_{rabbit}(t) - X_m(t)] - r_3[lb + r_4(ub - lb)], q < 0.5 \end{cases} \quad (3)$$

Where $X(t)$ and $X(t+1)$ represent the position of the individual at the current and next iteration, respectively; *t* is the iteration count; $X_{rand}(t)$ is a randomly

generated individual position; $X_{rabbit}(t)$ is the position of the prey; r_1, r_2, r_3, r_4 are random numbers between 0 and 1; q represents the random selection strategy, $X_m(t)$ and is the average position of the individuals.

(2) Evaluation. Harris's hawks observe the prey's movements to assess the best timing for attack. In the algorithm, this is achieved by calculating the positions of hawks around the prey (i.e., the current optimal solution), thereby determining whether to proceed with the attack (transition to the next phase) or to continue exploration. The escape energy formula is as follows:

$$E = 2E_0\left(1 - \frac{t}{T}\right) \tag{4}$$

Where E_0 represents the initial energy of the prey, a random number between [−1,1]; t is the iteration number, and T is the maximum number of iterations; The algorithm enters the search phase when $|E| \geq 1$, and it enters the exploitation phase when $|E| < 1$.

(3) Launching the Attack. During the attack phase, Harris's hawks adjust their capture strategy based on the escape energy of the prey (rabbit), evolving four specific attack modes:

When $0.5 \leq |E| < 1$and $r \geq 0.5$, the rabbit attempts to escape through random jumps. Harris's hawks will dive in a fast and sudden descent to catch the prey, using the soft besiege method. During this phase, the prey tries to escape in an irregular pattern, which is mathematically modeled as follows:

$$X(t+1) = \Delta X(t) - E|JX_{rabbit}(t) - X(t)| \tag{5}$$

Where $\Delta X(t)$ is the difference between the position of the rabbit and the position of the hawks, and J is a random jump strength coefficient between [0, 2] reflecting the prey's irregular escape pattern.

When $|E| < 0.5$ and $r \geq 0.5$, the rabbit's escape energy is insufficient. Harris's hawks directly adopt a forceful siege strategy:

$$X(t+1) = X_{rabbit}(t) - E|\Delta X(t)| \tag{6}$$

When $0.5 \leq |E| < 1$ and $r < 0.5$, the rabbit has a chance to escape and has sufficient escape energy. Harris's hawks employ a progressive rapid dive with a gentle

siege approach:

$$X(t+1)=\begin{cases}Y, f(Y)>f\big(X(t)\big)\\ Z, f(Z)>f\big(X(t)\big)\end{cases} \tag{7}$$

$$Y=X_{rabbit}(t)-E|JX_{rabbit}(t)-X(t)| \tag{8}$$

$$Z=Y+S*LF(2) \tag{9}$$

Where f（）represents the fitness function in Equation (7); Sdenotes a two-dimensional random vector in Equation (9), whose elements are random numbers between [0, 1], and L is the Lévy flight function.

When $|E|<0.5$ and $r<0.5$, the rabbit has a chance to escape but lacks sufficient escape energy. A direct aggressive siege would help the prey conserve energy and escape in the chaos. To ensure foolproof success, Harris's hawks form a robust encircling formation before the attack, gradually narrowing it down, employing a progressive rapid dive with a forceful siege:

$$X(t+1)=\begin{cases}Y, f(Y)>f\big(X(t)\big)\\ Z, f(Z)>f\big(X(t)\big)\end{cases} \tag{10}$$

$$Y=X_{rabbit}(t)-E|JX_{rabbit}(t)-X_m(t)| \tag{11}$$

$$Z=Y+S*LF(2) \tag{12}$$

Repeat steps (3) and (4); when the number of iterations of the algorithm reaches the maximum number of iterations, output the current position of the prey as the CA transition rule obtained from the mining process.

2.2.2 Mining the transition rules based on HHO

The HHO-based transformation rule mining algorithm consists of three modules: the data processing module, the main process module, and the core algorithm module.

(1) Data Processing Module:

In this module, multiple sampling points are generated based on land use and spatial variable data. These sampling points correspond to pixels on a raster map, with each pixel representing a cell. For each sampling point, attribute values and surrounding neighborhood states, such as distance to city center, freeways, and elevation, are calculated. Subsequently, all spatial variables are standardized to a range between 0 and

1 to facilitate comparison and computation, thereby preparing for subsequent optimization processes.

(2) Main Process Module:

During this stage, the sampling dataset is inputted, and initial parameters are set. The current state of each cell is then computed in preparation for mining transformation rules. Next, the transformation rules mined by the core optimization module are evaluated by calculating coverage and conversion rates. Rules that meet predefined thresholds (minimum coverage and minimum transition probability) are retained, while redundant elements in the transformation rules are removed through a rule pruning operator. Qualified rules are added to the rule list, and samples that satisfy the current rules are removed using a sample update operator. Finally, the final transformation rule list is outputted once the minimum sample threshold is reached.

(3) Core Optimizer Module:

In this module, a set of "Harris Hawks" is randomly generated using Equation (2), where each "hawk" is a multidimensional vector containing feature attributes (e.g., distance to freeways, number of urban cells in the neighborhood). By calculating the fitness value using Equation (13) in the current solution space, the best feature vectors in the current population and other feature vectors are randomly searched around. Each "hawk's" position is randomly updated according to Equation (3) to achieve a wide-range search for the optimal solution space. During the evaluation phase, the prey's escape energy E is calculated using Equation (4), and based on the escape energy, the decision is made whether to enter the local exploitation phase to update the positions. The new position's fitness value is then calculated and compared with the current position's fitness value to determine the update method for the latest vector. The optimal transformation rules are derived when the maximum number of iterations is reached, and these rules are then passed to the main process module for subsequent pruning steps.

Through these three stages, the Harris Hawks Optimization (HHO) algorithm can progressively optimize from initially randomly generated vectors, ultimately producing accurate and effective CA transition rules.

2.3 The Key Operators

2.3.1 Rule pruning

For all transition rules, further processing is required to eliminate redundant constraint attributes through rule pruning, aiming to enhance the predictive classification performance of the transition rules. Additionally, rule pruning helps prevent overfitting to training data. The basic idea behind rule pruning is to remove a constraint attribute, evaluate the quality of the pruned rule based on a fitness function, and only actually eliminate the attribute if the fitness value improves. This process is repeated, removing the next attribute until deleting any attribute would decrease the fitness value.

$$fitness = \left(\frac{TP}{TP \times FN}\right) \times \left(\frac{TN}{FP \times TN}\right) \tag{13}$$

To implement rule pruning, corresponding CA transition rule pruning operators need to be designed. In this paper, based on the characteristics of transition rules unearthed by the HHO algorithm, the implementation method of the pruning operator is represented in pseudocode as follows(see **Table 1**):

Table 1 Rule pruning operator pseudocode

Parameter description:
pre_rule: Transition rules before pruning
new_rule: Transition rules after pruning
sample_set: Training sample set
fields[i]: The i-th attribute condition in the transition rules
field_count: Total number of attribute conditions in the transition rules

Pseudocode：

```
Input sample_set;
Input pre_rule;
prefitness=fitness(pre_rule);
newfitness=prefitness;
do
    prefitness=newfitness;
    for i=0 to field_count;
        new_rule=PruningField(pre_rule,fields[i]);
        if fitness(new_rule)<fitness(pre_rule);
            BackPruningField(pre_rule,field[i]);
        else
            newfitness=fitness(new_rule);
```

13 **end for**
14 **while** newfitness⩾prefitness

2.3.2 Sample update

In order to minimize the occurrence of multiple transition rules operating on the same sample during the mining process and covering the original results, after each rule is discovered and pruned, the examples that already satisfy this rule can be removed from the sample set, thereby improving the quality of the rule set. Specifically, satisfying the condition means that the attributes of the example meet the rule, and the current cell status is the same as the predicted cell status.

2.4 The integrated HHO-CA model

This paper utilizes the transition rules discovered by the HHO algorithm to construct the HHO-CA model, which is then applied to the dynamic simulation of urban growth. The HHO-CA model consists of two main parts. The first part involves using HHO to mine the transition rules of CA, extracting these rules from spatial training data over two periods. The second part focuses on simulating urban growth using the obtained transition rules. In the simulation process, a sufficient number of iterations are necessary to ensure the generation of stable simulation results. In this experiment, 200 iterations are chosen. The basic structure and workflow of the HHO-CA model are illustrated in **Fig 2** below.

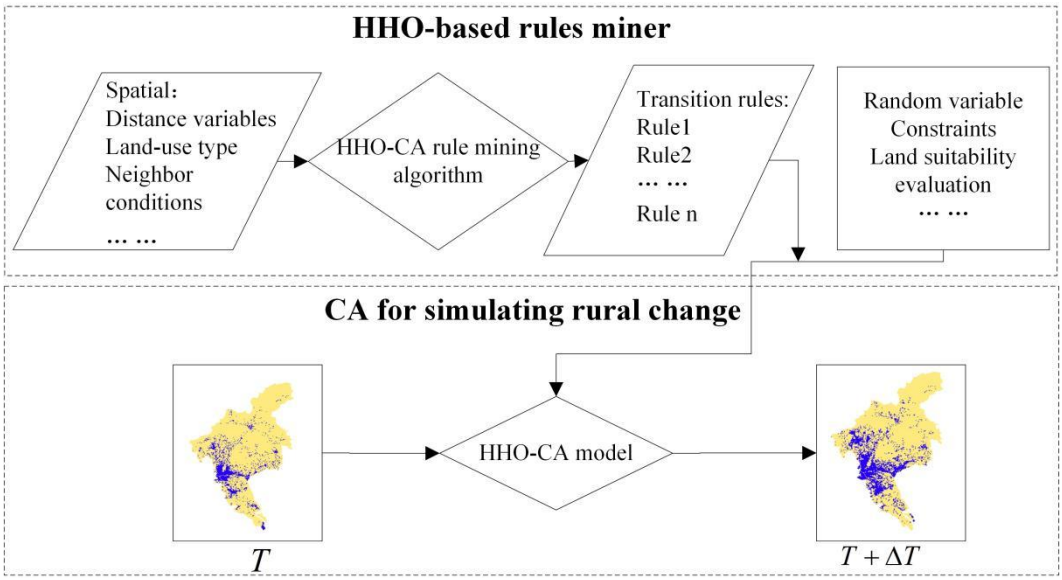

Fig 2. The basic structure and workflow of the HHO-CA model

3. Research area and model application

3.1 Area of study and experimental data

Guangzhou is located in the southern part of China, in the central and southern part of Guangdong Province, on the northern edge of the Pearl River Delta. The basic data selected for the experiment are land-use type data for Guangzhou in the years 2000, 2005, and 2010. Land-use types include arable land, forest land, grassland, water bodies, urban and rural construction land, and unused land. In these three periods of land-use data, the land-use data for 2000 and 2005 are used as training data for the mining of transition rules, while the land-use data for 2010 is used as test data to evaluate the quality of the model. Additionally, land-use types are reclassified into urban land and non-urban land, represented by 1 and 0, respectively.

In this paper, three main research variables are employed: distance variables, neighborhood conditions, and physical attributes (see **Fig 3**). The distance variables include distance to the city center (dis_city)(Ju et al., 2016), distance to the highway (dis_freeway)(Salem et al., 2018), distance to the railway (dis_railway), and distance to ordinary roads (dis_road). The neighborhood condition represents the normalized count of cells classified as urban within a 7×7 neighborhood window(neighbors). The physical attribute is elevation (DEM).This study utilized the Euclidean Distance tool in

the ArcGIS Spatial Analyst module to calculate the Euclidean distances for each distance variable on the raster maps. The data was then standardized uniformly using the Raster Calculator to facilitate subsequent data collection and rule extraction.

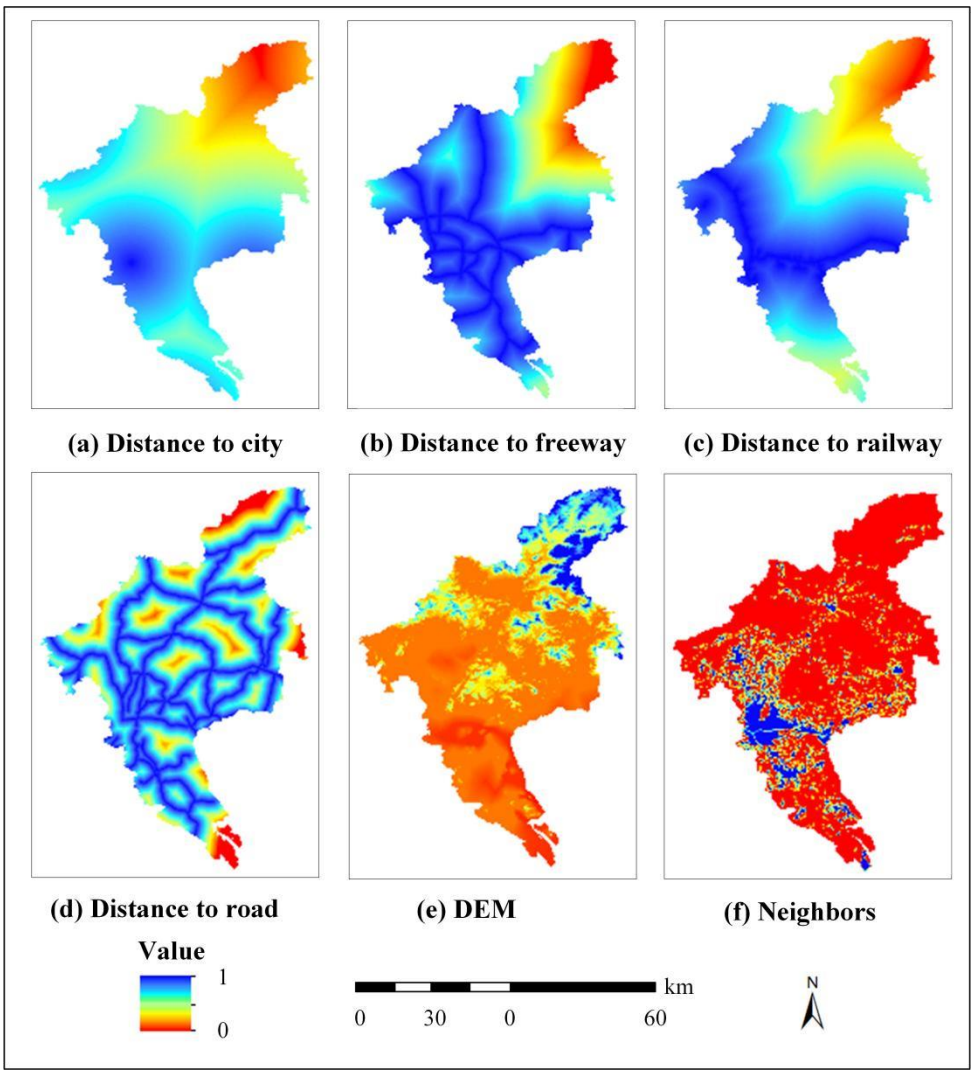

Fig 3. Visualization of some spatial variables

In this experimental process, a stratified random sampling method is used to extract samples from the training dataset. A total of 20,005 samples were randomly selected for mining transition rules (see **Fig 4**), and these samples were further divided into two groups: 15,000 samples as training data for constructing transition rules, and the remaining samples as test data for evaluating the quality of the transition rules.

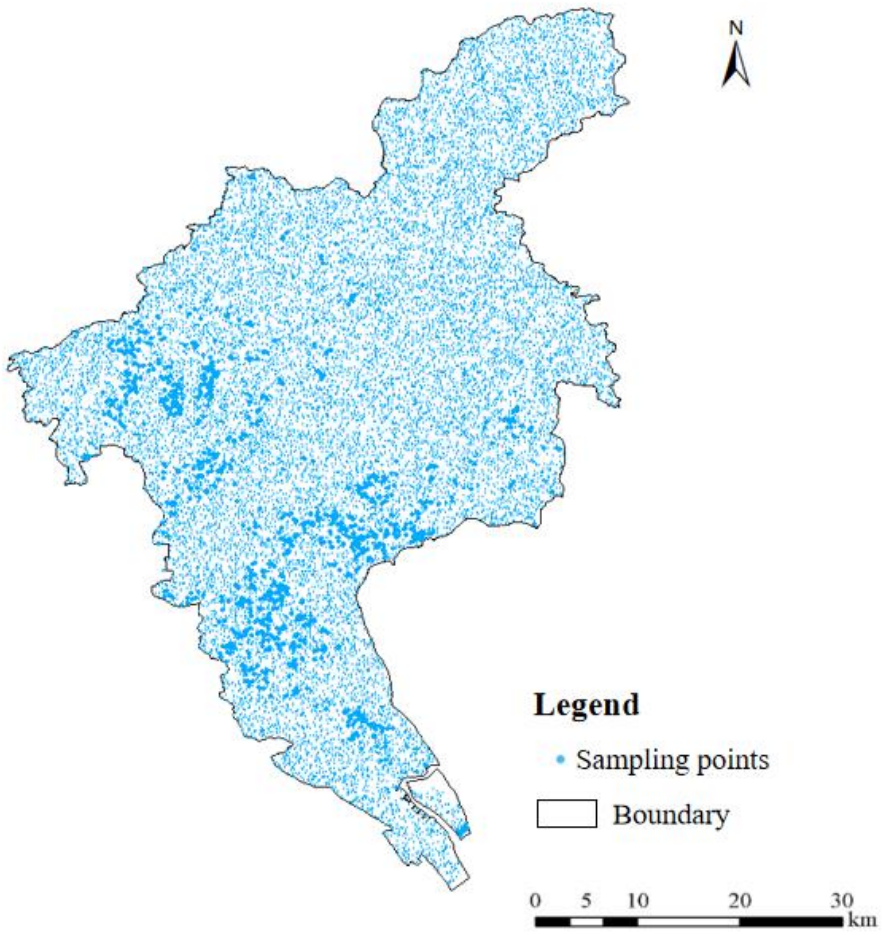

Fig 4. Distribution map of sampling points

3.2 Transition rules mining and urban simulation

After acquiring spatial data for the experimental area and setting various parameters in the algorithm, the parameters set in this study are as follows: N=30 represents the population size, i.e., the number of Harris Hawks, and T=500 represents the maximum number of iterations for mining transformation rules. Min_Coverage=0.001 indicates the minimum coverage rate, which is the minimum proportion of samples that a rule must cover during the rule mining process. Min_TransProb=0.7 denotes the minimum transition probability, indicating the proportion of cells that actually undergo the expected state change among those that satisfy a given rule. The predicted transition of cell status evolving into urban areas is designated as 1, otherwise as 0. The rules governing the simulation of urban growth in

the CA model are derived from training samples, and during the simulation process, 14 transition rules were identified. The **Table 2** below presents some of the transition rules along with their confidence levels and coverage rates.

Table 2 Some transition rules mined by the HHO algorithm

Rule 1:
If dis_freeway<=10204.571 and dis_railway>=157.202 and dis_road<=15152.817
Then class=non_urban(confidence=0.9896,coverage=0.0478)
Rule 2:
If dis_freeway>=8514.573 and dis_railway>=2893.047 and neighbors>=9
Then class=urban(confidence=0.8427,coverage=0.2218)
Rule 3:
If dis_freeway<=8936.542 and dis_railway>=143.215 and dis_road<=16215.721
Then class=non_urban(confidence=0.9124,coverage=0.0534)
Rule 4:
If dis_freeway>=8452.346 and dis_railway>=2789.127 and neighbors>=7
Then class=urban(confidence=0.8567,coverage=0.0021)
Rule 5:
If dis_city>=3640.710 and dis_freeway<=14766.847 and dis_railway<=4732.470 and dis_road>=12654.655 and dis_road<=17715.412 and dem>=8.544
Then class=non_urban(confidence=0.9891,coverage=0.0502)
Rule 6:
If dis_city<=7398.214 and dis_freeway<=9215.512 and dis_railway<=4312.857 and dis_road>=12346.532 and dem<=98.643
Then class=non_urban(confidence=0.9341,coverage=0.0678)
Rule 7:
If dis_city>=5643.210 and dis_freeway<=8112.214 and dis_railway>=3241.214 and dis_road<=15432.789 and neighbors>=14
Then class=urban(confidence=0.8734,coverage=0.0912)
Rule 8:
If dis_city>=2560.741 and dis_freeway<=10204.000 and dis_railway<=4912.300 and dis_road>=14254.671 and dis_road<=17415.712 and dem>=18.945
Then class=non_urban(confidence=0.9211,coverage=0.0486)
Rule 9:
If dis_freeway>=9021.243 and dis_railway<=2561.429 and dis_road<=16254.714 and neighbors>=5
Then class=urban(confidence=0.7993,coverage=0.0857)
Rule 10:
If dis_city<=6384.532 and dis_freeway<=9732.842 and dis_railway<=5234.728 and dis_road>=13456.215 and neighbors>=12
Then class=urban(confidence=0.8112,coverage=0.0925)
Rule 11:

If dis_city<=4500.321 and dis_freeway<=10012.325 and dis_railway<=3314.600 and dis_road>=15478.212 and dem<=129.214

Then class=non_urban(confidence=0.9225,coverage=0.0567)

Rule 12:

If dis_freeway<=8921.472 and dis_railway<=5123.612 and dis_road>=12367.214 and neighbors<=8

Then class=non_urban(confidence=0.9021,coverage=0.0612)

Rule 13:

If dis_city>=7341.214 and dis_freeway<=8143.215 and dis_railway<=4512.789 and dis_road>=14214.711 and dem>=45.214

Then class=non_urban(confidence=0.9134,coverage=0.0491)

Rule 14:

If dis_freeway>=16817.449 and dis_freeway<=17309.019 and dis_railway<=4892.829 and dis_road>=15611.212 and dis_road<=16840.963 and dem>=7.747 and neighbors>=3

Then class=urban(confidence=0.7778,coverage=0.0011)

In the **Table 2**, Rule 2 is interpreted to mean that a cell is likely to be converted into urban land if it meets the following conditions: the distance from the nearest freeway (dis_freeway) is equal to or greater than 8514.573, the distance from the nearest railway (dis_railway) is equal to or greater than 2893.047, and the number of neighbors is greater than or equal to 9. This scenario is associated with a confidence level of 0.8427 and encompasses 22.18% of the samples.

Applying the obtained transition rules to the extracted samples yielded an overall accuracy of 91.11%, with a Kappa coefficient of 0.8216, as detailed in **Table 3**.

Table 3 Accuracy of the samples

	Actual 2010 urban	Actual 2010 non-urban	Total
Simulation 2010 urban	10032	1624	11656
Simulation 2010 non-urban	155	8194	8349
Total	10187	9818	20005
Total accuracy	（10032+8194）/20005=91.11%		
Accuracy of urban cells	10032/10187=98.48%		
Kappa coefficient	0.8216		

The rules mined through the HHO algorithm were then employed in simulating the urban growth of Guangzhou from 2000 to 2010. The research assumes that the spatial variables and the relationship between land use changes during the experimental

process remain constant. The simulation process from 2000 to 2010 involved iterating the HHO-CA model 200 times, as illustrated in **Fig 5** below :

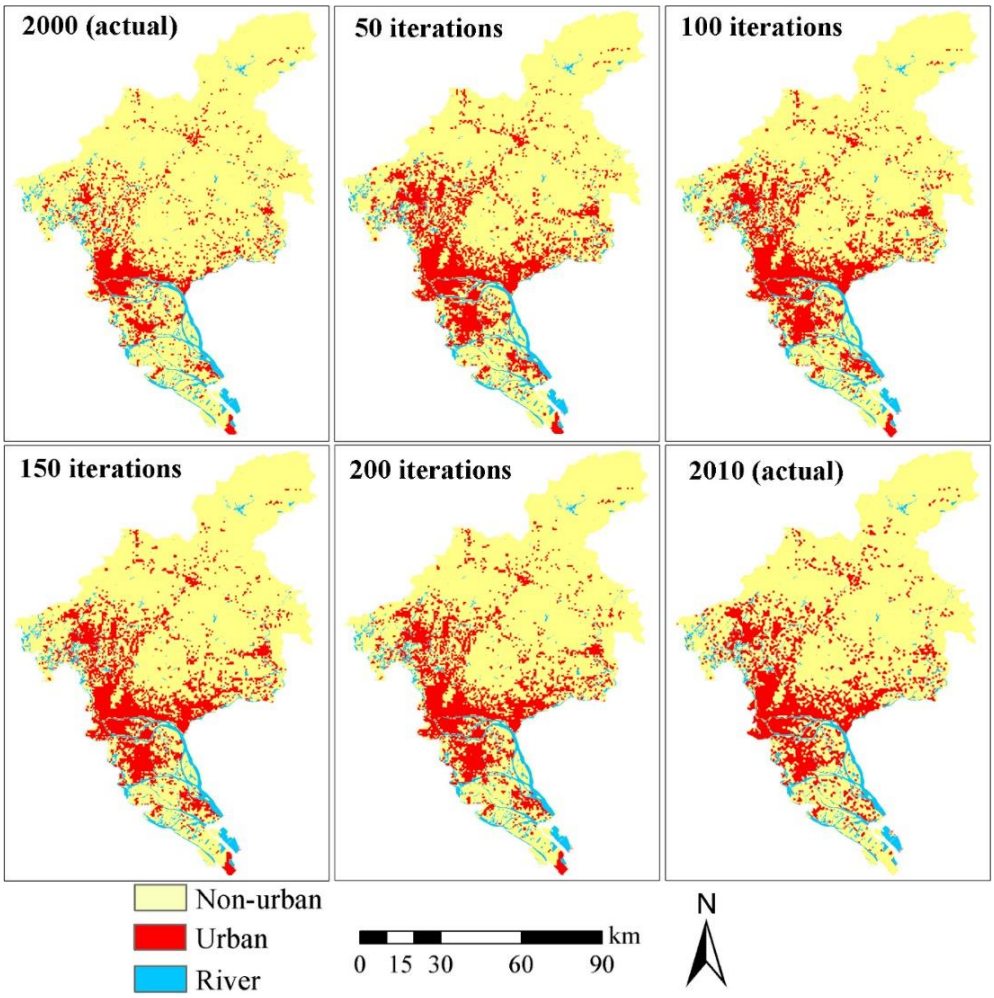

Fig 5.Iteration process diagram

4. Results and analysis

Model validation is crucial for assessing the accuracy of the CA model. The fundamental approach to CA model validation involves visually comparing simulated results with actual outcomes. Spatial overlay analysis was employed in this study to better illustrate the effectiveness of the model (see **Fig 6**). The comparison between simulated and real outcomes revealed a visually consistent alignment between the results obtained through the HHO-CA model and the actual data. To further accurately assess the transformation efficacy of the HHO-CA model, a unit-by-unit comparative approach was adopted, involving the computation of confusion matrices, overall

accuracy, and Kappa coefficients.

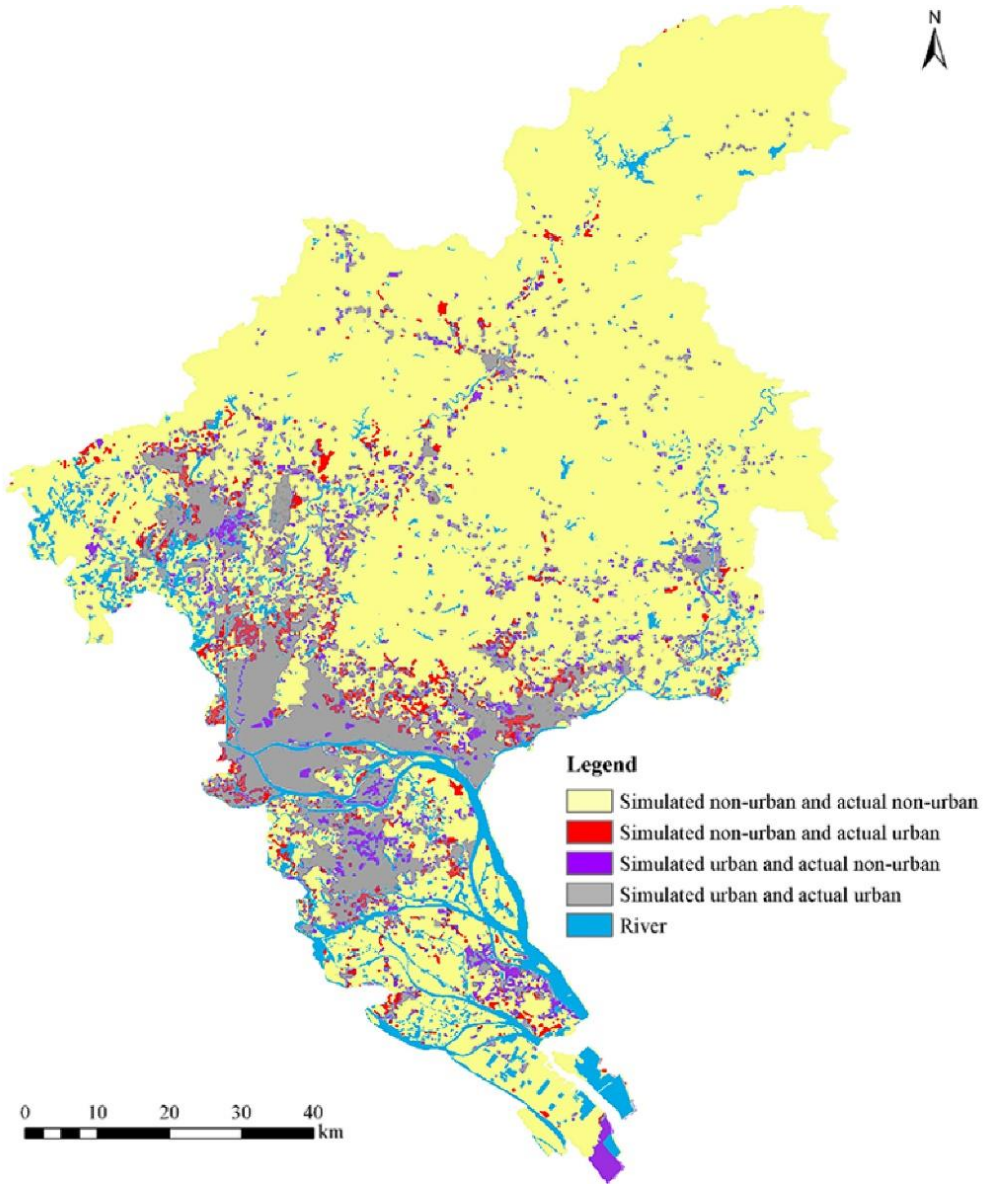

Fig 6. Spatial overlay diagram of HHO-CA model simulation

During the simulation process, 29,326 actual urban cells were classified as non-urban, while 42,062 actual non-urban cells were classified as urban. The overall accuracy was 91.95%, with a Kappa coefficient of 0.7441. The research findings suggest that commonly used prediction models exhibit better spatial consistency compared to the unchanged NULL model. The overall accuracy of the NULL model, calculated using land-use data from 2000 and 2010, was 90.57%, with a Kappa

coefficient of 0.6365 (see **Table 4**). Moreover, the overall accuracy of the simulation based on the HHO-CA model was 1.38% higher than that of the NULL model, with a Kappa coefficient of 0.1076 higher than that of the NULL model (see **Table 5**). This indicates that the HHO-CA model demonstrates excellent performance in practical applications, validating the effectiveness of the HHO algorithm in constructing the CA model, and making it suitable for modeling complex scenarios.

Table 4 Confusion matrix of the HHO-CA model simulation results

	Actual 2010 urban	Actual 2010 non-urban	Total
Simulation 2010 urban	**137552**	42062	179614
Simulation 2010 non-urban	29326	**678334**	707660
Total	166878	720396	**887274**
Total accuracy	（137552+678334）/887274=91.95%		
Accuracy of urban cells	137552/166878=82.43%		
Kappa coefficient	0.7441		

Table 5 Confusion matrix of the null model

	Actual 2010 urban	Actual 2010 non-urban	Total
Simulation 2010 urban	**92382**	9206	101588
Simulation 2010 non-urban	74496	**711190**	785686
Total	166878	720396	**887274**
Total accuracy	（92382+711190）/887274=90.57%		
Accuracy of urban cells	92382/166878=55.36%		
Kappa coefficient	0.6365		

Furthermore, spatial consistency is crucial for validating the CA model. Moran's I can be utilized to reflect the degree of spatial autocorrelation. Comparing the results simulated in 2010 with the actual data, the Moran's I of the HHO-CA model simulation in 2010 was found to be 0.685, while the actual Moran's I of 2010 was 0.719.

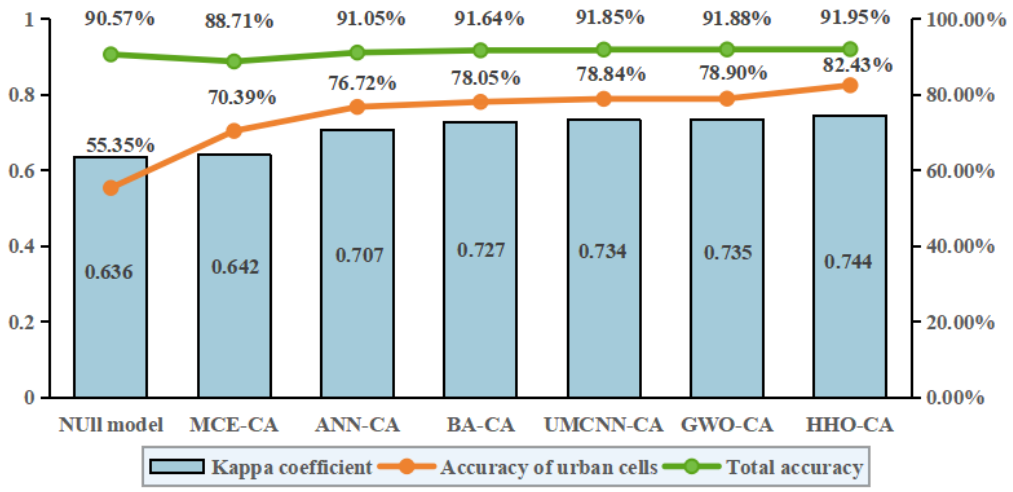

Fig 7. Comparisons of the Kappa coefficient, total accuracy, and accuracy of urban cell

To further explore the significant advantages of the HHO-CA model in simulating urban growth, a comparative analysis was conducted with other representative CA models. Common CA models such as MCE-CA, ANN-CA, BA-CA, UMCNN-CA, and GWO-CA were selected, all of which are well representative in simulating urban growth. These models were applied to the same dataset and spatial variables, and the results are shown in **Fig 7**(Ding et al., 2022). From the results, it can be observed that the HHO-CA model exhibits a significant improvement in the Kappa coefficient, total accuracy, and accuracy of urban cells. Compared to the second-best model, the HHO-CA model shows improvements of 0.009, 0.07%, and 3.53% in the Kappa coefficient, total accuracy, and accuracy of urban cells, respectively. Compared to the MCE-CA model, the improvements are 0.102, 3.24%, and 12.04% in the Kappa coefficient, total accuracy, and accuracy of urban cells, respectively. This indicates that the simulation results of the HHO-CA model are superior to other models, especially in terms of the Accuracy of urban cells, where the improvement is significant.

The validation and comparison of the aforementioned models demonstrate the outstanding performance of the HHO-CA model in simulating urban growth, showcasing remarkable performance throughout the entire simulation process. Compared to other optimization algorithms, the HHO algorithm exhibits significant advantages in mining the transition rules of CA models. This advantage stems from the HHO algorithm's effective exploration and utilization of the search space, as well as its

flexible adaptability to problem characteristics. This result not only strengthens confidence in the HHO-CA model for simulating urban growth but also provides valuable insights for further exploration and improvement of urban evolution research based on CA models.

5. Conclusion

The cellular automata (CA) model exhibits great potential in simulating the dynamic evolution of cities, but its accuracy largely depends on the accuracy of transition rules. This paper proposes a method to obtain geographic CA transition rules using the Harris Hawks optimization (HHO) algorithm, which successfully optimizes the upper and lower thresholds of each variable attribute by simulating the cooperative predation behavior of Harris's hawks. Experimental results demonstrate that HHO performs reliably and accurately in discovering CA transition models.

The study utilizes land-use data from Guangzhou, China, for the years 2000, 2005, and 2010, with data from 2000 and 2005 serving as training samples to simulate urban growth from 2000 to 2010. By evaluating the simulation results of 2010 against actual data, using the NULL model, and calculating overall accuracy and Kappa coefficient, the consistency between simulated results and actual data is verified. The results indicate that the accuracy of the HHO-CA model is significantly higher compared to the NULL model. Spatial pattern distributions using the HHO-CA model are analyzed through calculations of Moran’s I when compared to real data. Furthermore, compared to other classical CA models, the HHO-CA model exhibits significant improvements in the Kappa coefficient, total accuracy, and accuracy of urban cells, further illustrating its significant advantages in simulating urban growth.

This study further demonstrates that novel metaheuristic algorithms are suitable for defining transition rules in CA models and for studying the dynamic evolution of urban growth. The HHO-CA model can provide significant assistance in understanding urban evolution and urban spatial planning and can also offer decision support for urban planning and land use management.

However, there are still some areas for improvement in this research. For example,

the selection of variables for mining transition rules is not comprehensive enough. Urban growth is influenced not only by quantifiable factors such as economics and land use but also by a range of difficult-to-quantify social, cultural, and policy factors. Future research can attempt to comprehensively consider these non-quantifiable factors through qualitative analysis, case studies, and explore their correlation with urban growth patterns.

Future research will focus on the uncertain aspects of the model, considering more comprehensive factors affecting urban growth, thereby providing stronger evidence for using CA models for urban prediction. The impact of subtle numerical variations in CA conversion rules on model sensitivity will also be a significant issue for future investigation. Additionally, urban development is a multi-scale process, and future research will attempt to integrate data and models of different scales to more comprehensively understand the dynamic process of urban growth. Ultimately, the primary focus of this study is on the comparative analysis of transformation rule optimization algorithms. The incorporation of additional training data also presents a viable approach to enhance the simulation accuracy of the model. In future work, efforts will be directed towards exploring various avenues to further improve the model's simulation accuracy, adopting a more comprehensive perspective.

Declarations

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 42371397 and in part by the Marine Science and Technology Innovation Project of Jiangsu Province under Grant JSZRHYKJ202302.

Conflicts of Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Data availability

Data is provided within the manuscript or supplementary information files.

References

Batty M, Xie Y, Sun Z (1999) Modeling urban dynamics through GIS-based cellular automata. Computers, Environment and Urban Systems 23 (3):205-233. doi:https://doi.org/10.1016/S0198-9715(99)00015-0

Cai L, Wang M (2020) Effect of the thematic resolution of land use data on urban expansion simulations using the CA-Markov model. Arabian Journal of Geosciences 13 (23):1250. doi:10.1007/s12517-020-06248-z

Cao M, Tang GA, Shen QF, Wang YX (2015) A new discovery of transition rules for cellular automata by using cuckoo search algorithm. International Journal of Geographical Information Science 29 (5):806-824. doi:10.1080/13658816.2014.999245

Chen S, Feng Y, Tong X, Liu S, Xie H, Gao C, Lei Z (2020) Modeling ESV losses caused by urban expansion using cellular automata and geographically weighted regression. Science of The Total Environment 712:136509. doi:https://doi.org/10.1016/j.scitotenv.2020.136509

Clarke KC, Gaydos LJ (1998) Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. International Journal of Geographical Information Science 12 (7):699-714. doi:10.1080/136588198241617

Couclelis H (1997) From Cellular Automata to Urban Models: New Principles for Model Development and Implementation. Environment and Planning B: Planning and Design 24 (2):165-174. doi:10.1068/b240165

Ding Y, Cao K, Qiao W, Shao H, Yang Y, Li H (2022) A whale optimization algorithm–based cellular automata model for urban expansion simulation. International Journal of Applied Earth Observation and Geoinformation 115:103093. doi:https://doi.org/10.1016/j.jag.2022.103093

Guan Q, Shi X, Huang M, Lai C (2016) A hybrid parallel cellular automata model for urban growth simulation over GPU/CPU heterogeneous architectures. International Journal of Geographical Information Science 30 (3):494-514. doi:10.1080/13658816.2015.1039538

Hamdy O, Zhao S, Osman T, Salheen MA, Eid YY (2016) Applying a hybrid model of Markov chain and logistic regression to identify future urban sprawl in Abouelreesh, Aswan: A case study. Geosciences 6 (4):43

He C, Okada N, Zhang Q, Shi P, Li J (2008) Modelling dynamic urban expansion processes incorporating a potential model with cellular automata. Landscape and Urban Planning 86 (1):79-91. doi:https://doi.org/10.1016/j.landurbplan.2007.12.010

Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: Algorithm and applications. Future Generation Computer Systems 97:849-872. doi:https://doi.org/10.1016/j.future.2019.02.028

Huang Q, Liu Z, He C, Gou S, Bai Y, Wang Y, Shen M (2020) The occupation of cropland by global urban expansion from 1992 to 2016 and its implications. Environmental Research Letters 15 (8):084037

Huang X, Xu G, Xiao F (2021) Optimization of a Novel Urban Growth Simulation Model Integrating an Artificial Fish Swarm Algorithm and Cellular Automata for a Smart City. Sustainability 13 (4):2338

Ju H, Zhang Z, Zuo L, Wang J, Zhang S, Wang X, Zhao X (2016) Driving forces and their interactions of built-up land expansion based on the geographical detector–a case study of Beijing, China. International Journal of Geographical Information Science 30 (11):2188-2207

Karimi F, Sultana S, Babakan AS, Suthaharan S (2019) An enhanced support vector machine model for urban expansion prediction. Computers, Environment and Urban Systems 75:61-75

Li X, Liu X, Yu L (2014) A systematic sensitivity analysis of constrained cellular automata model for urban growth simulation based on different transition rules. International Journal of Geographical Information Science 28 (7):1317-1335. doi:10.1080/13658816.2014.883079

Li X, Yeh AG-O (2002) Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science 16 (4):323-343. doi:10.1080/13658810210137004

Liu XP, Li X, Liu L, He J, Ai B (2008) A bottom-up approach to discover transition rules of cellular automata using ant intelligence. International Journal of Geographical Information Science 22 (11-12):1247-1269. doi:10.1080/13658810701757510

Lu Y, Laffan S, Pettit C (2022) A geographically partitioned cellular automata model for the expansion of residential areas. Transactions in GIS 26 (3):1548-1571

Maria de Almeida C, Batty M, Vieira Monteiro AM, Câmara G, Soares-Filho BS, Cerqueira GC, Pennachin CL (2003) Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. Computers, Environment and Urban Systems 27 (5):481-509. doi:https://doi.org/10.1016/S0198-9715(02)00042-X

Msofe NK, Sheng L, Lyimo J (2019) Land Use Change Trends and Their Driving Forces in the Kilombero Valley Floodplain, Southeastern Tanzania. Sustainability 11 (2):505

Rubiera-Morollón F, Garrido-Yserte R (2020) Recent Literature about Urban Sprawl: A Renewed Relevance of the Phenomenon from the Perspective of Environmental Sustainability. Sustainability 12 (16):6551

Salem M, Bose A, Bashir B, Basak D, Roy S, Chowdhury IR, Alsalman A, Tsurusaki N (2021) Urban Expansion Simulation Based on Various Driving Factors Using a Logistic Regression Model: Delhi as a Case Study. Sustainability 13 (19):10805

Salem M, Tsurusaki N, Divigalpitiya P, Osman T Driving Factors of Urban Expansion in Peri-Urban Areas of Greater Cairo Region. In, 2018.

Song Y, Chen B, Kwan M-P (2020) How does urban expansion impact people's exposure to green environments? A comparative study of 290 Chinese cities. Journal of Cleaner Production 246:119018

Tayyebi A, Delavar MR, Yazdanpanah MJ, Pijanowski BC, Saeedi S, Tayyebi AH

(2010) A spatial logistic regression model for simulating land use patterns: a case study of the Shiraz Metropolitan area of Iran. Advances in earth observation of global change:27-42

Wei YD, Ewing R (2018) Urban expansion, sprawl and inequality. vol 177. Elsevier,

White R, Engelen G (1993) Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-Use Patterns. Environment and Planning A: Economy and Space 25 (8):1175-1199. doi:10.1068/a251175

Wu F (2002) Calibration of stochastic cellular automata: the application to rural-urban land conversions. International Journal of Geographical Information Science 16 (8):795-818. doi:10.1080/13658810210157769

Xie H, Zhang Y, Duan K (2020) Evolutionary overview of urban expansion based on bibliometric analysis in Web of Science from 1990 to 2019. Habitat International 95:102100. doi:https://doi.org/10.1016/j.habitatint.2019.102100

Zhong C, Guo H, Swan I, Gao P, Yao Q, Li H (2023) Evaluating trends, profits, and risks of global cities in recent urban expansion for advancing sustainable development. Habitat International 138:102869

Zhou Y, Wu T, Wang Y (2022) Urban expansion simulation and development-oriented zoning of rapidly urbanising areas: A case study of Hangzhou. Science of The Total Environment 807:150813. doi:https://doi.org/10.1016/j.scitotenv.2021.150813